



EUROPEAN EDUCATIONAL RESEARCH  
**Quality Indicators**



This project is funded by the Socioeconomic Sciences and Humanities Section.

**EERQI Final Conference, Brussels, 15-16 March 2011**

## **EERQI Basic Features**

- Part 1:** Ingrid Gogolin, University of Hamburg, EERQI  
Coordinator: Why EERQI, and what are EERQI's  
intentions?
- Part 2:** Stefan Gradmann, Humboldt University Berlin

**Prof. Dr. Stefan Gradmann**  
**Humboldt-Universität zu Berlin /**  
**Berlin School of Library and Information Science (IBI)**



# Overview

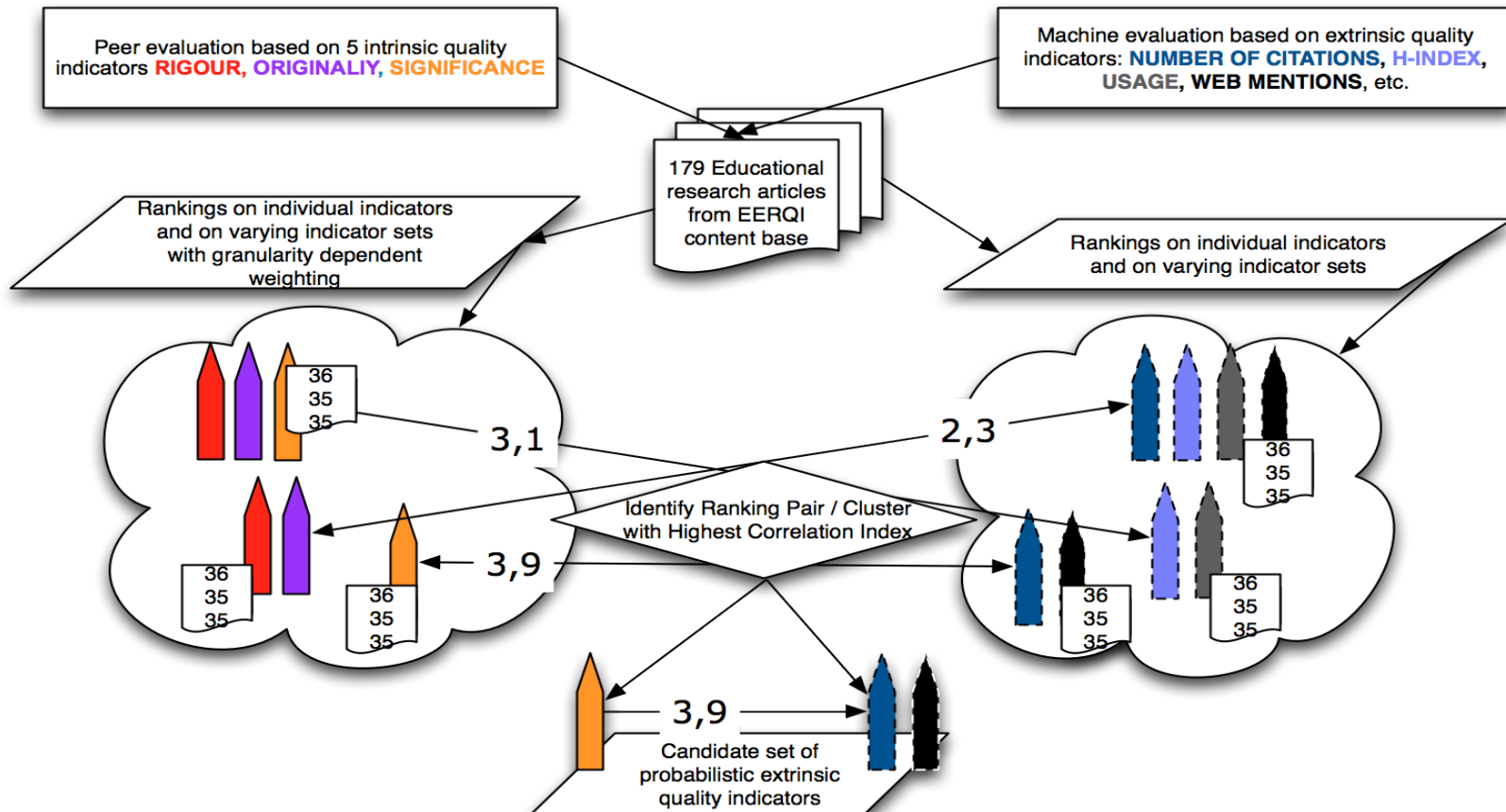
- **Transparency in Evaluation: an Issue of Data rather than of Methodology!**
- **... and Monographs?**

# Transparency: Source Data

- **~42.000 documents in the EERQI content base**
  - ~36.000 OA documents and ~6.000 documents from the publishing partners
- **Extrinsic Indicators**
  - Citation based measures such as, Scimago Journal Rank, Source Normalized Impact per Paper per author, h-index, g-index, e-index
    - Retrieved from SSCI, Scopus, Google Scholar
  - Usage indicators: based on data provided by the project MESUR and on usage data supplied by publishing partners
  - Web impact via mentions of author names (approach proposed by Mike Thellwall)
  - Social network services: mentioning of articles, authors or journals in sources such as Connotea, LibraryThing, CiteULike, Mendeley
- **A Tool for Indicator Retrieval: aMeasure**



# Transparency: Methodology



# Known and Newly Discovered Issues with Source Data

- **We were aware of unbalanced coverage in terms of**
  - scientific disciplines,
  - languages
  - publishing formats
- **... but only in the course of processing data we really got aware of intrinsic quality issues of our source data (documents and metadata):**
  - lack of document structures and of structural markup,
  - inconsistent character encoding,
  - lacking normalisation of entities,
  - lacking standardisation of attribute semantics
- **e. g. author identification and institutional or discipline affiliation, year of publication etc.**

## ... and Monographs

- **Monographs are still the dominating publishing format in large parts of the SSH**
  - Ratio of journals and monographs in our content base is roughly 1:3,5 (although this comparison already illustrates one of the methodological problems with monographs!)
- **Monographs (or: anything other than journal articles) are the 'terra incognita' of bibliometrics!**
  - How to weight?
  - What are typical time-frames of scholarly takeup?
  - How to count citations? What is a citation?? How to recognise one???
  - Which factors account for reputation? What about the publisher's brand?
  - And what about hybrid formats such as conference proceedings or edited volumes?

# Monographs: Emerging Approaches

- **Incomplete data coverage is the main problem**
- **The only viable data source is library catalogues**
- **Emerging approaches start from there:**
  - Libcitation (White et al., 2009)
    - Numerical monograph measure that „increases by one every time a different library reports acquiring that book in a national or international union catalog“
    - Assumed to cover all aspects of esteem, as well as the quality of publishers, because librarians jointly constitute a measuring instrument that is sensitive to them
    - Sample data with researcher publication data from 6 departments at two Australian universities, 2000-2006, retrieved manually
    - Lots of manual work needed due to data quality issues, once again
    - A viable measure for institutions and (with some restrictions) individual researchers.

# Monographs: Emerging Approaches

- **[library catalogues]**

- „Analogy Model“ (Torres-Salinas & Moed, 2009)
  - Catalogue data from 42 Spanish libraries retrieved using „ECONOM\*“ as search criterion via Z39.50 to gather a corpus of economics literature
  - application of traditional bibliometric indicators to their measurable counterparts in this corpus
  - Despite some limitations „the proposed analogy model between citation analysis of journal articles and library catalog analysis of book titles has proven to be valuable“.
- Linmans (2010) retrieved data relating to 80 authors 1,135 scholarly book titles, corresponding with 59,386 book holdings from WorldCat
  - Numerous problems in processing, mainly due to data inconsistency, lack of standardisation and hardly identifiable derivative relations!



## Work done in EERQI (XEROX)

- An attempt was done to apply citation detection methodology that had been used successfully in Computer Science to EERQI monographs
- The result is discouraging as in the following example:

“Mertens, Gerhard, Ursula Frost, and Winnfried Böhm. 2008. Handbuch der Erziehungswissenschaft.” is referred to as

- Mertens, Frost et al. 2008 – Handbuch der Erziehungswissenschaft
  - Mertens et al. (2008) Handbuch der Erziehungswissenschaft
  - Mertens, Frost, & Böhm, 2008
  - Mertens, Frost, & Böhm (2008)
  - Mertens, Frost and Böhm 2008
  - Mertens, Frost and Böhm (2008)
  - Mertens et al. (2008)
  - Mertens et al. 2008
- And we haven't yet addressed „ibid.“ or „a.a.o.“:)

## Work done in EERQI (Jenny Sieber)

- Attempt at processing PhD thesis comparing respective marks with bibliometric tracks in Google Books
  - Marks are not publicly available (and will not be so soon in the after-Guttenberg era :))
  - Citation frequency or numeric attributes of usage not recorded in Google Books
- Starting from White (2009) consider library union catalogs as a sorts of “librarians’ citation indexes.” Cases examined:
  - Karlsruhe Virtual Catalog (KVK) (Meta Search Engine)
  - Library Thing (Social Network)
  - WorldCat (Metacatalogue)
  - 40 monographs randomly selected from the EERQI content base and identified by their ISBN published between 2004 and 2010 in German (18), English (8) and French (14) were processed.

## Work done in EERQI (Jenny Sieber)

- **Some preliminary results:**
  - CatalogMentions show unexpected distributions, such as in the case of French monographs that are more prominent in WorldCat (100%) and even in the German National Library (86%) than in the French union catalogue ABES (79%)
  - Presence in Library Thing is still very low but easy to monitor using LT's API and an extension of our aMeasure tool → continue as an automated thread!
- **Very complex data and very complex objects: we've only just scratched the surface!**
- **A document covering our explorative study is available from Jenny Sieber.**

Thank you for your patience and attention!